

University of Zagreb

FACULTY OF CHEMICAL ENGINEERING AND TECHNOLOGY

Ivan Vrban

DATA-DRIVEN MODELLING AND MONITORING FOR ADVANCED CRYSTALLIZATION PROCESS CONTROL

DOCTORAL THESIS

Zagreb, 2026



University of Zagreb

FACULTY OF CHEMICAL ENGINEERING AND TECHNOLOGY

Ivan Vrban

DATA-DRIVEN MODELLING AND MONITORING FOR ADVANCED CRYSTALLIZATION PROCESS CONTROL

DOCTORAL THESIS

Mentor:

Prof. Nenad Bolf, PhD.

Zagreb, 2026



Sveučilište u Zagrebu

FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE

Ivan Vrban

MODELIRANJE I PRAĆENJE PROCESA KRISTALIZACIJE TEMELJENO NA PODATCIMA ZA NAPREDNO VOĐENJE PROCESA

DOKTORSKI RAD

Mentor:

Prof. dr. sc. Nenad Bolf

Zagreb, 2026.

ABSTRACT

Crystallization from solution is among the most widely used unit operations in pharmaceutical manufacturing, serving at once as a separation, a purification, and the route by which the solid-state properties of the active ingredient are established. It determines the polymorphic form, purity, crystal size and habit, and residual solvent level. It also governs downstream operations and, ultimately, the dissolution rate and bioavailability of the finished dosage form.

The outcome of crystallization depends on many interacting parameters, making the process complex and strongly non-linear. Reliable modelling, optimisation, and control are therefore essential to achieve consistent, predictable product properties. While mechanistic population balance models are the established framework for describing these dynamics, they are demanding to parameterise and rarely practical for routine process development and control.

At the same time, advances in process analytical technology have enabled in-situ monitoring of crystallization, producing large volumes of continuous process data. This data enables robust process control, data-driven modelling, and the subsequent optimisation of the crystallization process.

This dissertation develops and experimentally validates two distinct data-driven approaches built on this principle. The first provides real-time monitoring and control of supersaturation, the driving force of crystallization. The second provides predictive modelling and crystallization procedure optimisation.

The first approach develops a solute concentration sensor, based on spectroscopic and temperature data modelled with an artificial neural network. The sensor enables closed-loop supersaturation control that holds supersaturation at a given setpoint throughout the batch. It is demonstrated on a seeded cooling crystallization of ceritinib, where the network improved concentration prediction accuracy over the linear benchmark and supported stable closed-loop control of a live batch.

The second approach uses a long short-term memory recurrent neural network to predict the evolution of the crystal size distribution from routinely collected process variables. Feature engineering from the temperature profile substantially improves the accuracy of this predictive model. The trained model is then embedded as a digital twin inside a robustness-aware genetic algorithm. This algorithm identifies cooling trajectories and seeding conditions for prescribed target crystal sizes under realistic parametric uncertainty. Both stages are demonstrated on a

creatine monohydrate crystallization, where the optimised procedures were validated in laboratory batches.

The principal scientific contribution of this dissertation is the enforcement of robustness within the optimisation loop itself, through a data-driven digital twin embedded in a robustness-aware genetic algorithm. Taken together, the case studies show that data-driven models trained on experimental campaigns of realistic industrial size can support monitoring, control, and recipe design in pharmaceutical crystallization, using instrumentation that development laboratories already have.

Keywords: pharmaceutical crystallization, process analytical technology, artificial neural networks, long short-term memory, supersaturation control, robust optimisation, digital twin.

SAŽETAK

Kristalizacija iz otopine jedna je od najčešće primjenjivanih jediničnih operacija u farmaceutskoj proizvodnji jer istodobno služi kao postupak separacije i pročišćavanja. Tijekom kristalizacije određuju se polimorfni oblik, čistoća, veličina i habitus kristala te razina rezidualnih otapala. Također utječe na sve sljedne proizvodne operacije i na svojstva gotovog farmaceutskog oblika, kao što su brzina otapanja i bioraspoloživost.

Ishod procesa kristalizacije ovisi o velikom broju međusobno povezanih parametara, zbog čega je riječ o složenom i izrazito nelinearnom procesu. Pouzdano modeliranje, optimizacija i vođenje procesa ključni su za postizanje dosljednih i predvidljivih svojstava proizvoda. Modeli populacijske bilance predstavljaju ustaljeni mehanistički okvir za opisivanje dinamike kristalizacije. Određivanje vrijednosti parametara ovog tipa modela zahtjevno je i rijetko je praktično za rutinski razvoj procesa. Istodobno je razvoj procesno analitičkih tehnologija omogućio *in-situ* praćenje kristalizacije, pri čemu nastaju velike količine kontinuiranih procesnih podataka. Dobiveni podaci omogućuju učinkovito upravljanje procesom, razvoj modela temeljenih na podacima te optimizaciju procesa kristalizacije.

Disertacija razvija i eksperimentalno validira dva različita pristupa temeljena na podacima dobivenim procesno analitičkom tehnologijom. Prvi pristup omogućuje praćenje i vođenje prezasićenja u stvarnom vremenu, odnosno vođenje pokretačke sile kristalizacije. Drugi pristup omogućuje prediktivno modeliranje i optimizaciju postupka kristalizacije.

Prvi pristup razvija senzor koncentracije otopljene tvari temeljen na spektroskopskim i temperaturnim podacima modeliranima umjetnom neuronskom mrežom. Razvijeni senzor omogućuje vođenje prezasićenja, pri čemu se prezasićenje tijekom cijelog šaržnog procesa održava na zadanoj vrijednosti. Pristup je demonstriran na kristalizaciji ceritiniba hlađenjem, pri čemu je neuronska mreža poboljšala točnost predviđanja koncentracije u odnosu na linearni referentni model te omogućila stabilno vođenje prezasićenja tijekom procesa.

Drugi pristup koristi neuronsku mrežu s dugoročno kratkoročnim pamćenjem za predviđanje raspodjele veličine kristala na temelju rutinski prikupljenih procesnih varijabli. Inženjerstvo značajki izvedenih iz temperaturnog profila znatno povećava točnost prediktivnog modela. Razvijeni model zatim se ugrađuje kao digitalni blizanac u genetski algoritam koji uzima u obzir robusnost procesa. Algoritam određuje optimalne profile hlađenja i uvjete cijepljenja za postizanje zadanih veličina kristala uz parametarsku nesigurnost. Oba koraka

demonstrirana su na kristalizaciji kreatin monohidrata, pri čemu su optimirani postupci eksperimentalno validirani na laboratorijskoj skali.

Glavni znanstveni doprinos ove disertacije jest uvođenje robusnosti izravno u optimizacijsku petlju, primjenom digitalnog blizanca temeljenog na podacima ugrađenog u genetski algoritam koji uzima u obzir robusnost. Zajedno, prikazane studije slučaja pokazuju da modeli temeljeni na podacima, trenirani na eksperimentalnim skupovima podataka veličine usporedive s industrijskom praksom, mogu učinkovito poduprijeti praćenje, upravljanje i razvoj postupaka kristalizacije u farmaceutskoj proizvodnji, koristeći instrumentaciju koja je već dostupna u laboratorijima za razvoj procesa.

Ključne riječi: farmaceutska kristalizacija, procesna analitička tehnologija, umjetne neuronske mreže, neuronska mreža s dugoročno kratkoročnim pamćenjem, vođenje prezasićenja, robusna optimizacija, digitalni blizanac.